

CSIT5970 Advanced Cloud Computing
Midterm Examination Sample Questions

Question 1: Concept explanation

- (a) What is cloud computing?

- (b) What is virtualization?

- (c) Why does GFS/HDFS employ a large block?

Question 2: Multiple choices

- (1) The function of Secondary NameNode in HDFS is to
 - A. Serve as a backup for NameNode
 - B. Continue the functioning of NameNode
 - C. Serve as a checkpoint mechanism for primary NameNode
 - D. Provide advanced technology as compared with primary

- (2) If we increase the size of files stored in HDFS without increasing the number of files, then the memory required by NameNode
 - A. Decreases
 - B. Increases
 - C. Remains unchanged
 - D. Cannot be decided

Question 3: Pseudo-code programming

Write a MapReduce pseudo-code to count the occurrence of each word in a text file:

Question 4: Debugging

The following code is incorrect. Explain why and how it can be fixed.

```
1: class MAPPER
2:   method MAP(string  $t$ , integer  $r$ )
3:     EMIT(string  $t$ , integer  $r$ )

1: class COMBINER
2:   method COMBINE(string  $t$ , integers  $[r_1, r_2, \dots]$ )
3:      $sum \leftarrow 0$ 
4:      $cnt \leftarrow 0$ 
5:     for all integer  $r \in$  integers  $[r_1, r_2, \dots]$  do
6:        $sum \leftarrow sum + r$ 
7:        $cnt \leftarrow cnt + 1$ 
8:     EMIT(string  $t$ , pair ( $sum, cnt$ ))           ▷ Separate sum and count

1: class REDUCER
2:   method REDUCE(string  $t$ , pairs  $[(s_1, c_1), (s_2, c_2) \dots]$ )
3:      $sum \leftarrow 0$ 
4:      $cnt \leftarrow 0$ 
5:     for all pair  $(s, c) \in$  pairs  $[(s_1, c_1), (s_2, c_2) \dots]$  do
6:        $sum \leftarrow sum + s$ 
7:        $cnt \leftarrow cnt + c$ 
8:      $r_{avg} \leftarrow sum / cnt$ 
9:     EMIT(string  $t$ , integer  $r_{avg}$ )
```